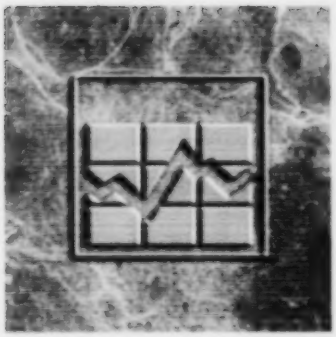
Catalogue no. 75F0002MIE — No. 002
ISSN: 1707-2840
ISBN: 978-0-662-47911-6

Research Paper

Income Research Paper Series

Survey of Labour and Income Dynamics: Possible Interview Dates



by Maryanne Webber

Income Statistics Canada
Jean Talon building, Ottawa, K1A 0T6

Telephone: 1-613-951-7355

Statistics Canada Statistique Canada

Canada

SURVEY OF LABOUR AND INCOME DYNAMICS: POSSIBLE INTERVIEW DATES

August 1992

Maryanne Webber, Household Surveys Division

The SLID Research Paper Series is intended to document detailed studies and important decisions for the Survey of Labour and Income Dynamics. These research papers are available in English and French, at no charge. To obtain a summary description of available documents or to obtain a copy of any, please contact Philip Giles, Manager, SLID Research Paper Series, by mail at 11-D8 Jean Talon Building, Statistics Canada, Ottawa, Ontario, CANADA K1A 0T6, by telephone (613) 951-2891, or by fax (613) 951-3253.

EXECUTIVE SUMMARY

When a survey respondent is asked to recall various events, it is known that the quality of the responses diminishes as the length of recall increases. On the other hand, increasing the frequency of data collection increases both the costs of collection and the burden on the respondents. The paper examines options which attempt to strike a reasonable balance between these factors. As it relates to this decision, the paper also describes how the sample has been designed to ensure that it remains representative of the target population, both for a given year and over time.

The conclusion is that, at this time, SLID should collect labour data in January to cover the previous calendar year and to collect income data in May, again to cover the previous calendar year.

TABLE OF CONTENTS

		Page
1.	INTRODUCTION	1
2.	SURVEY OBJECTIVES	1
3.	TWO VARIANTS ON OPTION 2	2
4.	OPTION 1 : INTERVIEWING IN JANUARY AND MAY	
	4.1. Strengths	5
	4.2. Weaknesses	6
5.	OPTION 2 : INTERVIEWING IN JANUARY AND JUNE	
	5.1. Strengths	8
	5.2. Weaknesses	9
6.	SELECTED OPTION	12
APPENDIX A:	MAINTAINING A REPRESENTATIVE SAMPLE	14

1. INTRODUCTION

Respondents to the Survey of Labour and Income Dynamics (SLID) will be interviewed twice a year. In the course of these two interviews labour and income information covering a full twelve-month period must be collected. There are several possible models with respect to when during the year interviews should be done, and what information should be collected in each interview. Two options were evaluated from the perspective of the survey's objectives, cost, quality and complexity:

Option 1 -- Collect labour information for a full calendar year in January
Collect income information for a full calendar year in May

Option 2 -- Split the collection of labour data into two interviews, each covering approximately six months
Collect income information for a full calendar year in May or June (i.e., near income tax time)

2. SURVEY OBJECTIVES

SLID is a household panel survey designed to measure changes over time in economic well-being, and to provide information on the determinants of such changes, particularly with reference to demographic, family and labour market events. The survey will focus on medium-term dynamics; the intention is to follow individuals for six years. With a six-year study period, most users are expected to focus on broad-brush behavioural patterns rather than on week-by-week details of how a person spends his or her time. This has implications for the level of detail of the information that SLID should attempt to collect. The level of detail sought and the length of the reference period must be considered in relation to each other.

An important principle emerging in the planning of SLID is that the survey is first and foremost a longitudinal survey. There is a requirement to produce what are referred to as annual cross-sectional data. Specifically, estimates of annual labour force activity weighted to the population in January of each year are sought for use in time series analysis. The requirement to produce these data is secondary, however, and when there is a conflict between the longitudinal and cross-sectional requirements, the former will be favoured. By the same token, the allocation of survey funds should be consistent with the priority placed on generating good longitudinal data.

Another important principle is that SLID should meet the needs of researchers in a range of disciplines, and should accordingly not be too narrowly focused in subject matter terms. An effort is being made to carry a reasonably broad range of variables, but limiting the depth or amount of detail collected on each topic. This approach is also consistent with the use of proxy reporting, which is dictated by the desire to maintain high response rates and to minimize respondent burden.

These general objectives and principles were borne in mind in the evaluation that follows.

3. TWO VARIANTS ON OPTION 2

In Option 1, data collection would take place in January (labour) and May (income). In Option 2, two sets of collection dates were considered. First, the possibility of collecting labour data in May and November, and of treating income as a supplement to the May interview was examined. This approach was judged problematic because of the differences in the labour and income reference periods. The non-alignment of the reference periods, as illustrated in Table 1, would create problems for the production of annual cross-sectional estimates, resulting in a

more complex design and higher collection costs. In particular, it would be necessary to trace and follow a large proportion of cohabitants who are no longer living with longitudinal respondents in order to produce the cross-sectional estimates.[1] This seems an inappropriate use of funds, since the survey's primary focus is on longitudinal data. Thus the methodology chosen should allow us to produce cross-sectional data as a (relatively low marginal cost) spin-off.

A second set of dates offered more promise: collect labour data in January and June, and treat income as a supplement to the June interview. The reference periods for labour would not be of exactly equal length (being 7 and 5 months respectively), but this seemed better than any option that provided two six-month reference periods. [2] Therefore this variant of Option 2 is compared to Option 1 in the evaluation that follows.

It should be noted that, in Option 1, any sponsored supplementary questions would be piggy-backed on the May income interview, which is relatively short. In Option 2, supplementary questions would be asked in January.

---

[1] Longitudinal respondents are people originally selected for the survey in January 1993. Cohabitants are people who move in with longitudinal respondents at a later date. The latter are included in the survey as long as they are living with a longitudinal respondent. This provides contextual household information on the longitudinal respondents. The cohabitants' information will also be used in producing annual cross-sectional estimates of labour force activity.

[2] A January-July approach is undesirable. After mid-June, the collection problems would become serious because of lost or misplaced tax information and, more importantly, higher non-response due to vacation. July is also a peak month for moving -- not an ideal time to interview because respondents in the middle of a move are more likely to be busy or distracted.

TABLE 1 - COMPARISON OF REFERENCE DATES AND INTERVIEW DATES

OPTION 1: Interviews in January and May

January: Labour data for previous calendar year
May: Income data for previous calendar year

OPTION 2A: Interviews in May and November

November: Labour data for previous six months
May: Labour data for previous six months and Income data for previous calendar year

OPTION 2B: Interviews in January and June

January: Labour data for previous seven months
June: Labour data for previous five months and Income data for previous calendar year

---

o - - - - o indicates the reference period "underline" indicates the interview months

4. OPTION 1 : INTERVIEWING IN JANUARY AND MAY

4.1. Strengths

The use of the calendar year as a reference period. The calendar year is a useful and meaningful framework for respondents. The linking of events to major holidays at the beginning of the calendar year is an aid to recall.

Simplicity. The two interviews use the same reference period. This approach is easy for users to understand. It requires less processing than an approach that splits labour into two interviews, and will therefore result in more timely data. People who have experience with other large longitudinal surveys urge us to keep the survey as simple as we can.

Content links between the labour and income interviews. Some consistency checks between the labour information and the income information are under consideration, for example, a range check that compares earnings reported during the labour interview with the wages and salaries figure collected during the income interview. Another example: information on the receipt (yes / no) of UI will be collected in the labour interview, but amounts received will be reported in the income interview. These links are complex, and they are less viable if labour is collected at two time points. If checks between the labour and income data cannot be achieved in the field, the result will be additional post-collection processing (imputation rather than collecting data directly from the respondent).

4.2. Weaknesses

Twelve-month recall period. The main issues here are labour market spell data, complex labour market patterns and the reporting of social assistance and UI receipt. With respect to labour, straightforward situations (full-time, full-year workers with one job, retired persons, etc), which represent the majority of the population, are not problematic. Recall is a concern with more complex work patterns involving, for example, spells of unemployment and inactivity, variable work schedules, and so on. These unusual patterns (which are analytically interesting) may be under-reported.

The magnitude of errors can be anticipated to some extent from comparisons that have been done between Labour Force Survey (LFS) and Labour Market Activity Survey (LMAS) data. The general pattern tends to be an under-reporting of activity in LMAS data at the beginning of the year, coupled with some over-reporting towards the end of the year. LMAS employment estimates for 1986 were 1.2% below the LFS estimates in the early part of the year and 1.2% higher by the end of the year. With respect to unemployment, various definitions are possible with LMAS. The definition that provides the best comparison to LFS yields unemployment estimates of 8.5% under at the beginning of the year and about 6.9% over at the end of the year. These results (from the first year of LMAS) do not have the benefit of any feeding back of information from the previous year -- they represent the worse-case scenario for SLID.

There is some evidence that feeding back will improve the quality of reporting significantly. LMAS fed back one piece of information, employer name. A study of job separations shows that feeding back this information served to essentially eliminate the December-January seam; LMAS results tracked separations data derived from the Record of Employment file (an administrative

file maintained by Employment and Immigration Canada) and the LFS quite well. In addition, feeding back the employer name appears to have caused a drop in the number of people recorded as not having worked all year.

Concern about recall error must also be placed in the context of expected data uses. SLID expects to produce six years of data on each respondent. Even with a twelve month reference period, we expect to identify spells of employment quite accurately. Intermittent workers and other groups of interest will, in a six year file, be identifiable from their flow into and out of employment -- there should be no problem isolating such target populations. SLID will contribute to the understanding of labour market mechanisms through its relatively long reference period, rather than through the refinement of detail. Thus the intention is not to replicate the LFS for a period of six years, but rather to provide a broader perspective.

Another concern with the twelve month reference period relates to subannual income data. We are now looking at what will be possible in this area, and it appears rather limited. The main objective with respect to income is good annual data, and we will attempt to achieve this objective by moving closer to the information required on the Income Tax T1 form.[3] The collection or derivation of subannual data is a secondary objective. In the case of wages and salaries, it may be possible to derive monthly totals. For self-employment income, accurate monthly data is not really feasible (and its usefulness, even if feasible, is unclear). The meaning and usefulness of investment income on a subannual basis is also

---

[3] The Survey of Consumer Finances asks respondents to use their tax records for the reporting of certain items. SLID intends is to push this practice further. There is some evidence from focus-group testing that respondents find it easier to transcribe the information from tax records than to provide estimates from recall. The coverage of tax records is also very high. Indeed, we would like to offer respondents the choice of providing their income information by interview, or authorizing access to tax records. (This approach will be tested in 1994.)

doubtful. Unemployment Insurance (UI) and Social Assistance (SA) receipt are a special concern. The annual UI amount is reported on the T1 form, but obviously not the months in which UI was received. These months will be identified in the course of the labour interview. Recall may be a constraint here, but recall is not the only problem in reporting UI; sensitivity is also an issue. The same comments apply to SA data, with the added complication that SA is not currently reported on the T1 form. This will change for the 1992 tax year, but the quality of annual SA data remains a concern because the amounts are not presently provided to income tax filers on information slips[4]. Again, it is possible that collecting the data at two occasions during the year would result in higher quality, but quality is also thought to be affected also by an unwillingness to report the receipt of SA.

Turnaround time. A six-week period is anticipated for the labour interviews. This leaves about two months to turn around the information in preparation for the income interview. Given the need to adjust interviewer assignments (to reflect moves, sample attrition), this is quite tight.

5. OPTION 2 : INTERVIEWING IN JANUARY AND JUNE

5.1. Strengths

Shorter recall period for labour questions. The most significant gains in quality would relate to information on non-working periods. As noted above, unemployment may be underestimated early in the year when a one-year reference period is used. A shorter reference period would perhaps (a) improve the accuracy of data on transitions between unemployment and inactivity, and (b) increase the number of short non-working spells reported early in the year. This implies that

---

[4] To date, only Ontario and Manitoba have made a commitment to provide such information slips.

the start and end dates of jobs would be more accurately reported. A shorter recall period may also improve the quality of some job characteristics, for example, promotions, employer-sponsored training, and wages (particularly tips and commissions).

Shorter recall period for reporting the receipt of UI and Social Assistance. The receipt of UI and SA will, in all probability, be identified during the labour interview. (This makes sense because UI and SA are often associated with periods of not working.) Conducting the labour interview at two occasions would probably reduce recall error in the UI and SA reporting. The magnitude of the gains realized would depend on the causes of under-reporting of these income sources.

Better information on changes in household composition. The SLID approach to capturing changes in household composition will be to establish a roster of household members at the time of each interview, and to collect arrival and departure dates for persons joining and leaving the household. Information on household dynamics will be best if the two interviews are about six months apart. Otherwise, the quality of reporting move dates, and our success in tracing, will probably vary with the length of the interval between interviews.

5.2. Weaknesses

More seams. Cutting the reference period in half means doubling the opportunities for discontinuities at the seams; i.e. between reference periods. The result will be more processing, more difficulties in matching up jobs and in making the links between labour and income.

Impact on interviewing costs and respondent burden. For people with relatively straightforward annual work patterns, the collection of labour

information at two occasions in the year involves an unnecessary doubling of costs and burden. The additional cost has been estimated at close to $80,000 per year. The additional burden could have an effect on sample attrition which cannot be estimated. This approach would also entail higher interviewer training costs (interviewers will require training prior to every SLID labour interview). This increase is estimated at $90,000.

Additional complexity in the field. The labour module in January would cover the previous seven months and the one in June would cover the previous five months. In addition, the interviewer needs the capacity to conduct a twelve-month labour interview in January for people who became household members between June and January. Each CAI (computer-assisted interviewing) version of the labour questionnaire would contain a unique set of question wordings and date edits. This is not an insurmountable problem but programming effort and disk space required for the extra set of questions would reduce the capacity for designing and implementing other field edits. The SLID schedule is quite tight and resources are stretched. Additional complexities add to the risk that critical deadlines will not be met, or that important mistakes will occur.

Impact on processing costs and timeliness. Impacts include longer development time (since the processing system would be more complex) and longer processing time, in order to link labour data collected twice as often. Processing costs for labour would be effectively doubled. Release dates are likely to be set back by about three months, partly because of increased complexity, partly because the income information would be collected in June rather than May.

Higher non-response. If the interviews begin in early June, the interviewers have only two weeks in which to complete tracing and collection. The third week in June is reserved for LFS. Beyond that, vacations and moves will hamper data

collection (particularly in Quebec, where moves are highly concentrated in the period between St-Jean Baptiste Day (June 24) and the July long weekend). If the SLID cut-off is before the LFS Survey Week, response rates will suffer. If interviewing continues after the LFS Survey Week, the response gains will be small in relation to interviewing expenses.

Start-up problems. With this option, a labour interview should be conducted in June 1993. However, the CAI infrastructure will not be in place at that time, and funds do not exist for an additional interview in the 1992-93 fiscal year. In fact, if we were to adopt this approach, the only realistic course would be to introduce the mid-year labour interview in 1994 rather than 1993, as follows:

- January 1993: preliminary interview
- January 1994: 1993 labour interview (and supplement, if applicable)
- June 1994: 1993 income interview and January-May labour interview
- January 1995: Jun-Dec labour interview and supplement

Thus CAI would not be used until January 1994. This approach is reasonably close to what was originally planned, which is important in view of the short development time available.

6. SELECTED OPTION

Despite the potential gains in terms of respondent recall, and thus in terms of data quality, the SLID team on the whole prefers Option 1 -- conducting a labour interview in January and an income interview in May, for reasons of cost, timeliness, response burden and simplicity. However, there are clearly benefits to the January-June option, and the choice is not a easy one. The study of the two options served as an opportunity to spell out a strategy for coping with the limitations imposed by the one-year reference period. The main elements of this strategy are as follows:

Structure the interview in a way that improves recall. Improvements in interviewing style have been introduced in the successive annual labour surveys conducted by Statistics Canada (Annual Work Patterns Survey, Survey of Work History, Labour Market Activity Survey). The accumulated experience from these surveys shows that reporting is better when respondents are asked first to recall events that are relatively easy to remember (eg, when a job started) and then asked to "hinge" other events to the ones that have already been reported (eg, job search prior to the job).

Use dependent interviewing. On all the major variables, the plan is to use dependent interviewing to dampen seam effects. This will make it possible to provide spell data for periods that extend over the seam.

Assess recall in the dress rehearsal. In 1993, SLID will conduct a dress rehearsal on a sample of 2,800 households. This sample will be selected from respondents to the LFS who rotated out in May 1992. Thus, there are five months of LFS data already available on these respondents (January to May 1992). In January 1993, we will ask these respondents to report their labour market activities for the whole of 1992. Evaluation of the dress rehearsal results will include a

comparison of results obtained retrospectively to those originally reported to the LFS. Although comparisons will be hampered to some extent by questionnaire differences, useful results are anticipated. The objective is not only to study recall error but to place it in the context of response errors in general, including proxy effect, unavailability of the information requested, sensitivity of the information sought and so on. While it will be difficult to get a very good handle on all of this, the dress rehearsal data will no doubt improve our understanding of sources of error.

APPENDIX A:
MAINTAINING A REPRESENTATIVE SAMPLE

The target population for SLID is all persons living in Canada, excluding persons in the Yukon and Northwest Territories, residents of institutions (staying at least six months), persons living on Reserves, and full-time members of the Canadian Armed Forces.

According to the recommended option in this report, a sample of 20,000 households (one panel) will be selected every three years. At the time of sample selection, each panel will be representative of the target population. Each panel will remain in the SLID sample for six years. The first panel will be introduced in January 1993, the second in January 1996, and so on. Table 2 below illustrates the panel introduction, data collection and panel rotation cycles.

It is important to note that the SLID sample will be all persons in sampled households when the panel is introduced, and not the dwelling unit itself, which is often used as the sampling unit when a sample is selected from an area frame. *All sampled persons will remain in the sample for the life of the panel, regardless of whether any or all move*. At the time of any SLID interview, anyone who has moved in with a person in the SLID longitudinal sample (called "cohabitants") will also be interviewed. The reasons for this approach will be explained below.

Interviews will be conducted twice each year: in January for labour market data and in May for income data. In both cases, the reference period will be the previous calendar year. The persons to be interviewed for labour and income data are all those 15 years or older as of January 1 of the reference year (and therefore 16 years or older on January 1 of the year of data collection). However, all persons in the originally selected households, regardless of age, are deemed to be

in the target population, and are therefore followed if they move. Basic demographic information will be collected on all household members at each interview, even though some persons will be too young for the SLID interviews.

Each year the composition of the target population will change; thus changes to the sample are required to keep the sample representative. There are three possible sources of additions to the target population:

a) persons arriving from another country, from the Yukon or Northwest Territories, or from a Reserve;

b) former residents of institutions;

c) newborns.

Persons in these groups who "move in" with a person who is part of the SLID longitudinal sample will be included, since all cohabitants are also included in the survey. (Cohabitants are not considered part of the SLID longitudinal sample, and therefore are interviewed only as long as they reside with a member of the SLID longitudinal sample.) Therefore, the only source of bias from changes in the target population comes from new households where *all members were outside the target population at the time the sample was selected*. Based on information from other surveys, the number of households in this group is likely to be small, resulting in a small bias only. The possibility of sampling households in this group is currently being examined, but it is currently felt that this bias will not seriously affect the quality of the estimates, and thus will be ignored.

Reductions in the target population result from similar, but opposite, reasons as additions:

a) persons moving to another country, to the Yukon or Northwest Territories, or to a Reserve;

b) persons entering institutions;

c) deaths.

Although they will be out-of-scope for that particular SLID interview, an attempt will be made to trace all persons in groups (a) and (b) as they may re-enter the target population at a later point in time. Most deaths will be accounted for during the tracing process.

The implementation of this survey plan will ensure a representative sample both for a particular year as well as over time. For longitudinal analysis, the target population is defined as of January 1 of the year in which the panel is introduced. For cross-sectional analysis, the target population is defined as of January 1 of the reference year under analysis.

One potentially major source of bias will be non-response and sample attrition. The extended period of time during which a sampled person will be interviewed for SLID will result in the loss of respondents at every interview. Experience will tell whether the six-year time frame is unrealistic. If sample attrition becomes so high that data quality is seriously affected, a reduction in panel rotation will be considered. By the same token, if a high response rate can be maintained, the possibility of longer panels will also be considered.

TABLE 2

SAMPLE ROTATION PLAN														
	Year													
	93	94	95	96	97	98	99	00	01	02	03	04	05	06
Panel														
1	P	I	I	I	I	I	I							
2				P	I	I	I	I	I	I				
3							P	I	I	I	I	I	I	
4										P	I	I	I	I
P = Preliminary Interview I = Labour and Income Interviews														
Each panel has 20,000 households at the outset.														